ΠΜΣ ΣΤΗ ΧΡΗΜΑΤΟΟΙΚΟΝΟΜΙΚΗ ΤΕΧΝΟΛΟΓΙΑ
Τμήμα Λογιστικής και Χρηματοοικονομικής
Πανεπιστήμιο Μακεδονίας

ΕΝΤΥΠΟ ΥΠΟΒΟΛΗΣ ΠΡΟΤΑΣΗΣ ΘΕΜΑΤΟΣ ΔΙΠΛΩΜΑΤΙΚΗΣ ΕΡΓΑΣΙΑΣ

(για ένταξη στην Τράπεζα Θεμάτων Διπλωματικών Εργασιών του Π.Μ.Σ.)

1. ΣΤΟΙΧΕΙΑ ΠΡΟΤΑΣΗΣ

Πεδίο	Στοιχεία
Κωδικός Θέματος *(συμπληρώνεται από τη Γραμματεία μετά την έγκριση του θέματος από τη Συντονιστική Επιτροπή)*	
Ημερομηνία Υποβολής	28/06/2026
Προτείνων	Καθηγητής Νικόλαος Σαμαράς
Φορέας Προέλευσης Θέματος *(FAC, FTSAI, RES, IND, STU, EXT)*[1]	FAC
Κύρια Θεματική Περιοχή *(FINTECH, , RISK, AI-DATA,, DLT, REG, GOV, PROG, IND)*	AI-DATA
Δευτερεύουσα Θεματική Περιοχή *(προαιρετικά)*	FINTECH
Τριτεύουσα Θεματική Περιοχή *(προαιρετικά)*	--

[1] **FAC:** Μέλος Δ.Ε.Π. ή Διδάσκων του Π.Μ.Σ., **FTSAI:** Financial Technology and Strategic Artificial Intelligence Laboratory, **RES:** Άλλη ερευνητική δομή ή ερευνητικό έργο, **IND:** Επιχείρηση ή οργανισμός, **STU:** Πρόταση φοιτητή, **EXT:** Εξωτερικός συνεργάτης ή φορέας.

2. ΤΙΤΛΟΣ ΘΕΜΑΤΟΣ

Τίτλος στα Ελληνικά

Υπολογιστική σύγκριση αλγορίθμων συσταδοποίησης

Title in English

Clustering Algorithms Performance

Acronym: **CLAP** — **CL**ustering **A**lgorithms **P**erformance

3. ΚΑΤΗΓΟΡΙΑ ΔΙΠΛΩΜΑΤΙΚΗΣ ΕΡΓΑΣΙΑΣ

- ☐ Ερευνητική Διπλωματική Εργασία
- ☒ Εφαρμοσμένη Διπλωματική Εργασία
- ☒ Τεχνολογική Διπλωματική Εργασία
- ☐ Διπλωματική σε Συνεργασία με Οργανισμό ή Επιχείρηση
- ☐ Διπλωματική Ενταγμένη σε Ερευνητική Δράση

Εφόσον επιλεγεί η τελευταία κατηγορία — Όνομα Ερευνητικής Δράσης

4. ΜΕΘΟΔΟΛΟΓΙΚΗ ΠΡΟΣΕΓΓΙΣΗ

- ☒ Βιβλιογραφική ή Θεωρητική Μελέτη
- ☐ Συστηματική Βιβλιογραφική Ανασκόπηση
- ☐ Εμπειρική ή Ποσοτική Ανάλυση
- ☐ Μελέτη Περίπτωσης
- ☒ Συγκριτική Ανάλυση
- ☐ Ανάπτυξη ή Αξιολόγηση Τεχνολογικού Συστήματος
- ☒ Σχεδιασμός Πλαισίου, Μεθοδολογίας ή Μοντέλου
- ☐ Μικτή Προσέγγιση

5. ΣΥΝΤΟΜΗ ΠΕΡΙΓΡΑΦΗ ΤΟΥ ΘΕΜΑΤΟΣ

(ενδεικτική έκταση: 100 έως 250 λέξεις)

Η συσταδοποίηση (clustering) αποτελεί θεμελιώδη τεχνική μη επιβλεπόμενης/καθοδηγούμενης μάθησης, με στόχο την ανακάλυψη φυσικής δομής σε δεδομένα χωρίς εκ των προτέρων γνωστές ετικέτες: τα αντικείμενα ομαδοποιούνται ώστε όσα ανήκουν στην ίδια συστάδα να εμφανίζουν υψηλή ομοιότητα και όσα ανήκουν σε διαφορετικές να διαφέρουν. Παρά την ευρεία εφαρμογή της σε πεδία όπως η εξόρυξη δεδομένων, η βιοπληροφορική και η ανάλυση εικόνας, δεν υπάρχει ενιαίος αλγόριθμος βέλτιστος για κάθε περίπτωση. Κάθε οικογένεια αλγορίθμων στηρίζεται σε διαφορετικές παραδοχές για τη μορφή, το πλήθος και την πυκνότητα των συστάδων.

Η παρούσα διπλωματική εργασία πραγματοποιεί συστηματική υπολογιστική σύγκριση αντιπροσωπευτικών αλγορίθμων συσταδοποίησης από διαφορετικές κατηγορίες: διαμεριστικών (k-means, k-medoids), ιεραρχικών, βασισμένων στην πυκνότητα (DBSCAN, OPTICS), φασματικών και πιθανοτικών (μείγματα Gaussian). Η αποτίμηση στηρίζεται σε πολλαπλούς άξονες — στην ποιότητα των συστάδων μέσω εσωτερικών (silhouette, Davies–Bouldin, Calinski–Harabasz) και εξωτερικών (ARI, NMI) μετρικών, στο υπολογιστικό κόστος και την κλιμάκωση ως προς το πλήθος δειγμάτων και διαστάσεων, καθώς και στην ευρωστία απέναντι σε θόρυβο, ακραίες τιμές και επιλογή υπερπαραμέτρων.

Τα πειράματα διεξάγονται σε συνθετικά και πραγματικά σύνολα δεδομένων με ποικίλα χαρακτηριστικά, ώστε να αναδειχθούν τα συγκριτικά πλεονεκτήματα και οι περιορισμοί κάθε προσέγγισης. Απώτερος στόχος είναι η εξαγωγή τεκμηριωμένων συμπερασμάτων και πρακτικών κατευθύνσεων για την επιλογή κατάλληλου αλγορίθμου ανάλογα με τη φύση των δεδομένων και τους περιορισμούς της εκάστοτε εφαρμογής.

6. ΣΤΟΧΟΙ ΚΑΙ ΕΡΕΥΝΗΤΙΚΑ ΕΡΩΤΗΜΑΤΑ

Α. Στόχοι

- Υλοποίηση ενός αναπαραγώγιμου πειραματικού πλαισίου σύγκρισης, που να επιτρέπει την αποτίμηση αντιπροσωπευτικών αλγορίθμων συσταδοποίησης από διαφορετικές κατηγορίες (διαμεριστικών, ιεραρχικών, βασισμένων στην πυκνότητα, φασματικών, πιθανοτικών) υπό κοινές και ελεγχόμενες συνθήκες.
- Εξαγωγή κανόνων για την επιλογή κατάλληλου αλγορίθμου συσταδοποίησης ανάλογα με τα χαρακτηριστικά των δεδομένων (πλήθος εγγραφών και

διαστάσεων (χαρακτηριστικών), γεωμετρία και πυκνότητα συστάδων, επίπεδο θορύβου) και τους περιορισμούς της εκάστοτε εφαρμογής.

Β. Ερευνητικά Ερωτήματα

- Πώς μεταβάλλεται η σχετική ποιότητα των συστάδων που παράγουν οι διάφορες οικογένειες αλγορίθμων ανάλογα με τη φύση των δεδομένων (σφαιρικές έναντι μη κυρτών συστάδων, ισορροπημένα έναντι ανισομεγέθη μεγέθη, παρουσία θορύβου), και σε ποιο βαθμό συμφωνούν μεταξύ τους οι εσωτερικές και οι εξωτερικές μετρικές αποτίμησης;
- Πόσο ευαίσθητος είναι κάθε αλγόριθμος στην επιλογή των υπερπαραμέτρων του (π.χ. πλήθος συστάδων, παράμετροι πυκνότητας) και στην παρουσία θορύβου ή ακραίων τιμών, και ποιες μέθοδοι αναδεικνύονται πιο εύρωστες και αξιόπιστες σε μη ιδανικές συνθήκες;

7. ΠΡΟΤΕΙΝΟΜΕΝΗ ΜΕΘΟΔΟΛΟΓΙΑ

Η μεθοδολογία που θα ακολουθηθεί είναι η εξής:

1. Εκμάθηση όλων των βιβλιοθηκών και των εντολών της Python σχετικά με τεχνικές συσταδοποίησης.
2. Σχεδιασμός μελέτης και επιλογή αλγορίθμων
3. Επιλογή data sets από το UC Irvine Machine Learning Repository και Kaggle
4. Σχεδιασμός πλαισίου αξιολόγησης (μετρικές)
5. Εκτέλεση υπολογιστικής σύγκρισης στα επιλεγμένα data sets
6. Ανάλυση και παρουσίαση των αποτελεσμάτων

8. ΔΕΔΟΜΕΝΑ ΚΑΙ ΠΗΓΕΣ ΔΕΔΟΜΕΝΩΝ

https://www.kaggle.com/

https://archive.ics.uci.edu/

https://www.kaggle.com/datasets?tags=11108-Finance

https://finance.yahoo.com/

Βιβλιογραφία:

- Hans Petter Langtangen, A Primer on Scientific Programming with Python
- Graham J.Williams Simeon J. Simoff (Eds.), Data Mining: Theory, Methodology, Techniques, and Applications
- Blanco-Silva et al, Learning SciPy for Numerical and Scientific Computing
- Jiawei Han,Micheline Kamber ,Data Mining : Concepts and Techniques
- Ian H.Witten, Eibe Frank, Data Mining: Practical Machine Learning Tools and Techniques

9. ΠΡΟΑΠΑΙΤΟΥΜΕΝΕΣ ΓΝΩΣΕΙΣ Ή ΔΕΞΙΟΤΗΤΕΣ

A. Hard skills

- Βασικές Γνώσεις Στατιστικής
- Γνώσεις αλγορίθμων συσταδοποίησης
- Γνώσεις Προγραμματισμού σε Python
- Συγγραφή κειμένου σε LaTeX

B. Soft skills

- Σύνταξη κατάλληλων prompts σε AI tools
- Ορθή διαχείριση χρόνου
- Ανάληψη πρωτοβουλιών
- Κριτική σκέψη

10. ΑΝΑΜΕΝΟΜΕΝΑ ΠΑΡΑΔΟΤΕΑ

Κείμενο διπλωματικής εργασίας
Το πειραματικό πλαίσιο σύγκρισης
Οδηγό επιλογής αλγορίθμου συσταδοποίησης ανάλογα με τα δεδομένα
Πιθανή επιστημονική δημοσίευση σε επιστημονικό Conference/Journal

11. ΕΚΤΙΜΗΣΗ ΔΥΣΚΟΛΙΑΣ

Κλίμακα 1 (χαμηλή) έως 5 (υψηλή)

Κατηγορία	1	2	3	4	5
Θεωρητική Δυσκολία	☐	☐	☒	☐	☐
Προγραμματιστική Δυσκολία	☐	☐	☐	☐	☒
Δυσκολία Συλλογής Δεδομένων	☐	☐	☐	☒	☐

12. ΕΚΤΙΜΗΣΗ ΔΥΣΚΟΛΙΑΣ

☐ Χαμηλή

☒ Μέτρια

☐ Υψηλή

Σύντομη αιτιολόγηση

Ο χώρος είναι ώριμος και ανταγωνιστικός. Ο μεταπτυχιακός φοιτητής οφείλει να τοποθετηθεί στο όχι «πώς θα το φτιάξω», αλλά «πώς θα το κάνω να αξίζει σε ένα πεδίο που έχουν ήδη ειπωθεί πολλά». Η υλοποίηση είναι μέτριας δυσκολίας αλλά η **πειστική συνεισφορά και η μεθοδολογική εντιμότητα** είναι η πρόκληση.

Συντελεστές που *μειώνουν* τη δυσκολία: Πλήρης διαθεσιμότητα εργαλείων στο scikit-learn, τυποποιημένη στατιστική μεθοδολογία, άφθονη βιβλιογραφία ως οδηγός.

Συντελεστές που *αυξάνουν* τη δυσκολία: Η ανάγκη πρωτότυπης οπτικής σε ένα ήδη κορεσμένο πεδίο, η διασφάλιση δίκαιης σύγκρισης των αλγορίθμων συσταδοποίησης, η πειραματική αυστηρότητα (seeds, ομοιόμορφο πρωτόκολλο) και η φιλοδοξία τα συμπεράσματα να είναι γενικεύσιμα και όχι ad hoc.